Am. J. Hum. Genet. 50:1231–1234, 1992

Linkage of Familial Breast Cancer to Chromosome 17q21 May Not Be Restricted to Early-Onset Disease

Patricia Margaritte,* Catherine Bonaiti-Pellie,* Mary-Claire King,† and Françoise Clerget-Darpoux*

*Unité de Recherches d'Epidémiologie Génétique, INSERM Unité 155, Château de Longchamp, Paris; and †School of Public Health and Department of Molecular and Cell Biology, University of California, Berkeley

Summary

Lod scores for linkage between familial breast and ovarian cancer and markers on chromosome 17q21 are more frequently positive among families with disease diagnosed at younger ages than they are among older-onset families, suggesting that linkage is restricted to early-onset disease. However, for late-onset cases, the relative probability of sporadic rather than inherited disease is higher than previously suggested. If this correction is made, then later-onset families are much less informative; linkage heterogeneity based on age at onset is no longer significant; and for the sample of families as a whole, linkage is significant at a recombination fraction since demonstrated to be close to the correct locale. There is probably more than one gene for inherited breast cancer, but heterogeneity may not be due to age at disease onset.

Introduction

In some families, breast cancer susceptibility is inherited as an autosomal dominant trait (Williams and Anderson 1984; Newman et al. 1988; Claus et al. 1991; Iselius et al. 1991). However, because sporadic (i.e., noninherited) breast cancer is common and the disease allele rare, sporadic cases may appear in families with inherited disease. Furthermore, inherited breast cancer may be heterogeneous—i.e., due to mutations at different loci in different families. Linkage analysis can be a suitable approach for identifying such loci by locating them on the genome.

Analysis of extended families with multiple cases of breast cancer (and ovarian cancer in some families) indicated linkage of breast and ovarian cancer to chromosome 17q21 in some families (Hall et al. 1990; Narod et al. 1991). For the first series of 23 families analyzed, lod scores for linkage between disease and the most polymorphic marker D17S74 were inversely correlated with the mean ages at breast cancer diagnosis (Hall et al. 1990). At recombination fraction (θ) .01, the cumulative lod score for the seven families in which average age at diagnosis is ≤ 45 years was 5.98. The cumulative lod score for the 14 families in which average age at diagnosis is <52 years remained positive, but the cumulative lod score for all 23 families combined was -5.48 at close linkage to D17S74. Homogeneity could be rejected. The lod score adjusted for heterogeneity was 3.28 at $\theta = .014$ from D17S74, with disease linked to this locus in 40% of the families. Failure to adjust for heterogeneity led to a lower maximum lod score (2.35) at a higher θ (.20) from D17S74. More recent evidence indicates that this breast and ovarian cancer gene is located near D17S579, at about $\theta = .16$ proximal to D17S74 (Hall et al. 1992).

The 23 extended families described above were selected for multiple cases of breast cancer in order to maximize the likelihood that a disease gene was segregating in each. However, because sporadic breast cancer is so frequent, even multiply affected families may include one or more sporadic cases. Also, because risk of sporadic breast cancer increases more rapidly with age than does risk of inherited breast cancer, families with older-onset disease are more likely to include sporadic cases. In this paper, we suggest that linkage het-

Received September 12, 1991; revision received January 10, 1992.

Address for correspondence and reprints: Françoise Clerget-Darpoux, INSERM U. 155, Château de Longchamp, Bois de Boulogne, 75016 Paris, France.

0002-9297/92/5006-0010$02.00

erogeneity dependent on age at onset may have been an artifact of underestimating the relative probability of sporadic disease among cases with late age at onset.

Methods

If A is the disease allele with frequency q, then among affected individuals, genotypes AA and Aa represent inherited cases and genotype aa represents sporadic cases. The relative risk of being a sporadic case is specified by the probabilities of genotypes AA, Aa, or aa of pedigree members, given their phenotypes. Phenotype information includes disease status, age at diagnosis, and sex for affected relatives and sex and current age for unaffected relatives. For an affected individual, the probability of genotype y, given that disease occurred in age interval x, is deduced from the risk $i(y,x)$ that an individual with genotype y developed the disease exactly in age interval x—i.e., the density function. For an unaffected individual, the probability of genotype y, given no disease by age x, is deduced from the probability of remaining disease free until x, given y—i.e., $I - I(y,x)$ where $I(y,x)$ is the cumulative risk of disease up to age x, given genotype y. The relative probability R_x of sporadic versus inherited disease for an affected individual of age x is computed from the probabilities of each genotype that are deduced from the cumulative incidence function $I(y,x)$:

$$R_x = [(1-q)^2 I(y = aa,x)]/\{[q^2 + 2q(1-q)]\, I(y = AA \text{ or } Aa,x)\}.$$

The relative probability R'_x of sporadic versus inherited disease for an affected individual diagnosed at age x is computed from the genotype probabilities deduced from the incidence function $i(y,x)$:

$$R'_x = [(1-q)^2 i(y = aa,x)]/\{[q^2 + 2q(1-q)]\, i(y = AA \text{ or } Aa,x)\}.$$

Linkage analysis was carried out using the MLINK program from LINKAGE (Lathrop et al. 1984), with liability classes defined by incidence risks for affected individuals and by cumulative risks for unaffected individuals. Homogeneity of linkage was evaluated by the admixture test (Smith 1963) by using the program HOMOG (Ott 1991), by the B-test (Risch 1988), and by the "predivided sample test" (Morton 1956).

Results

For cumulative incidence $I(y,x)$ and incidence $i(y,x)$, table 1 indicates the relative probabilities R_x and R'_x of sporadic versus inherited disease for a randomly selected breast cancer patient. The relative probabilities R_x reflect the previous analysis by Hall et al. (1990), in which risks for unaffected individuals were defined by $I - I(y,x)$, as above, and in which risks for affected individuals were also estimated from cumulative risks $I(y,x)$—i.e., .37 by age 40 years, .66 by age 55 years, and .82 over the entire lifetime, for genotypes AA and Aa; and .004 by age 40 years, .028 by age 55 years, and .081 over the entire lifetime, for genotype aa. The relative probabilities R'_x reflect estimates from incidence risks for unaffected individuals, as recommended above.

As expected, R_x and R'_x are equal for the lowest-age class but diverge at higher-age classes. R' is twice R for the interval 40–55 years and is more than threefold larger for ages above 55 years. For affected individuals, estimating risk from the cumulative risk $I(y,x)$ rather than from the risk $i(y,x)$ leads to underestimation of the probability of being a sporadic case, in the older-age intervals.

Table 1

Risk of Breast Cancer, Given Genotype at a Susceptibility Locus, and Relative Probability *R* of Sporadic versus Inherited Breast Cancer, for Affected Women

Age (years)	Risk Defined by Cumulative Incidence $I(y, x)$			Risk Defined by Density Function $i(y, x)$		
	y = AA or Aa	y = aa	R_x	y = AA or Aa	y = aa	R_x
15–39.....	.37	.004	.53	.37	.004	.53
40–54.....	.66	.028	2.08	.29	.024	4.06
≥55.....	.82	.081	4.84	.16	.053	16.23

Lod scores for linkage of breast cancer to D17S74 calculated at two θ's by using the present model are indicated in table 2, with the previously published results shown for comparison. There are two striking differences. First, the maximum lod score, 3.52 at θ = .12, is significant for the sample of families as a whole. Second, evidence for heterogeneity disappears: X_1^2 = 1.63 for the admixture test; X_1^2 = 1.31 for the B-test, and X_1^2 = 2.34 for a predivided-sample test, when families 1–7 are compared with families 8–23.

Families in which breast cancer appears in both women and men (families 16 and 19) have substantially negative lod scores under both models. If these families are compared with families that have only affected women, heterogeneity is significant on the basis of the predivided-sample test: X_1^2 = 5.16, $P < .05$. When the two families with male breast cancer are excluded, the maximum lod score for the remaining 21 families is 4.64 at θ = .08 from D17S74.

Discussion

The a priori probability that an affected woman from the general population will be a sporadic case was underestimated in the oldest-age intervals in the previous analysis of these families. Consequently, the a posteriori probability, given the family history of a late-onset case being sporadic, was underestimated as well, although to a lesser extent. Sporadic cases have probability .5 of recombining with a marker. Hence, if sporadic cases were misclassified as inherited cases, then apparent recombination events would lead to very negative lod scores and apparent heterogeneity. When older-onset cases were given a higher probability of being sporadic, then lod scores for families with late-onset disease were close to zero and evidence for heterogeneity was reduced.

Lod scores are still higher for families with early-onset disease than for families with late-onset disease, probably because of the differential informativeness of families. In families with early-onset disease, affected

Table 2

Lod Scores from Linkage Analysis of 23 Breast Cancer Famlies with D17S74 at Two θ's

Family	Average Age at Diagnosis[a] (years)	Lod Score in: Hall et al. (1990); θ = .001	Lod Score in: Present Study; θ = .001	Lod Score in: Present Study; θ_{max} = .12
1	32.7	2.36	2.25	1.69
2	37.2	.50	.75	.50
3	37.3	.40	.43	.29
4	39.8	1.14	1.36	1.03
5	42.6	−.50	−.68	−.14
6	44.2	1.38	1.22	.85
7	45.4	.70	.90	.65
8	47.0	.00	−.08	−.03
9	47.4	−.31	−.07	.16
10	47.6	−.04	−.03	−.03
11	49.3	−1.51	−.71	−.22
12	50.2	−.06	−.21	−.10
13	50.4	−.41	−.43	.00
14	51.4	−.65	.05	.15
15	51.8	−.35	−.36	−.02
16	52.0	−2.71	−1.97	−.38
17	53.5	−.13	.12	.14
18	53.6	−.75	−.19	−.07
19	55.8	−2.56	−1.90	−.59
20	56.4	−1.71	−.98	−.71
21	58.7	.65	.41	.27
22	59.4	−.85	−.06	.06
23	63.3	−.07	.03	.02
Total		−5.48	−.15	3.52

[a] Of breast cancer in the family.

individuals have a high probability of being inherited cases, so linkage is informative. In contrast, families with older mean age at diagnosis may include several cases with a nonnegligible probability of being sporadic, so there is little linkage information. Thus, lod scores for older-onset families are close to zero. The consistently negative lod scores for families with affected men, as well as the significant heterogeneity between these families and the other families, suggest that familial male breast cancer probably has another etiology.

Linkage heterogeneity for inherited breast cancer is not excluded, but it may not be related to age at disease onset. In particular, the breast cancer gene on chromosome 17q21 may influence both early- and late-onset disease. The locale for the breast cancer gene suggested by this reanalysis of the original data is closer to the critical region revealed by direct mapping studies (Hall et al. 1992). This analysis also shows that liability classes should be defined differently for affected relatives than for unaffected relatives. More generally, as has been discussed elsewhere (Bishop et al. 1988; Narod and Amos 1990; Skolnick et al. 1990; Claus et al. 1991; Iselius et al. 1991), linkage analysis is sensitive to the specification of liability classes, particularly for diseases with a high frequency of sporadic cases.

References

Bishop DT, Cannon-Albright L, McLellan T, Gardner EJ. Skolnick MH (1988) Segregation and linkage analysis of nine Utah breast cancer pedigrees. Genet Epidemiol 5: 151–169

Claus EB, Risch N, Thompson WD (1991) Genetic analysis of breast cancer in the cancer and steroid hormone study. Am J Hum Genet 48:232–242

Hall JM, Friedman L, Guenther C, Lee MK, Weber JL, Black DM, King M-C (1992) Closing in on a breast cancer gene on chromosome 17q. Am J Hum Genet 50:1235–1242

Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B, King M-C (1990) Linkage of early-onset familial breast cancer to chromosome 17q21. Science 250:1684–1689

Iselius L, Slack J, Littler M, Morton NE (1991) Genetic epidemiology of breast cancer in Britain. Ann Hum Genet 55:151–159

Lathrop GM, Lalouel J-M, Julier C, Ott J (1984) Strategies for multilocus linkage analysis in humans. Proc Natl Acad Sci USA 81:3443–3446

Morton NE (1956) The detection and estimation of linkage between the genes for elliptocytosis and the Rh blood type. Am J Hum Genet 8:80–96

Narod SA, Amos C (1990) Estimating the power of linkage analysis in hereditary breast cancer. Am J Hum Genet 46: 266–272

Narod SA, Feunteun J, Lynch HT, Watson P, Conway T, Lynch J, Lenoir GM (1991) Familial breast-ovarian cancer locus on chromosome 17q12-q23. Lancet 338:82–83

Newman B, Austin MA, Lee M, King M-C (1988) Inheritance of human breast cancer: evidence for autosomal dominant transmission in high risk families. Proc Natl Acad Sci USA 85:3044–3048

Ott J (1991) Analysis of human genetic linkage, rev. ed. Johns Hopkins University Press, Baltimore

Risch N (1988) A new statistical test for linkage heterogeneity. Am J Hum Genet 42:353–364

Skolnick MH, Cannon-Albright LA, Goldgar DE, Ward JH, Marshall JC, Schuman BG, Hogle H, et al (1990) Inheritance of proliferative breast disease in breast cancer kindreds. Science 250:1715–1720

Smith CAB (1963) Testing for heterogeneity of recombination values in human genetics. Ann Hum Genet 27:175–182

Williams WR, Anderson DE (1984) Genetic epidemiology of breast cancer: segregation analysis of 200 Danish pedigrees. Genet Epidemiol 1:7–20